NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

A COMPLEX ADAPTIVE SYSTEM APPROACH TO FORECASTING HURRICANE TRACKS

by

Matthew R. Lear

June 2005

Thesis Advisor: John Hiles
Second Reader: Patrick Harr

Approved for public release; distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			*Form Approved OMB No. 0704-0188*
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington DC 20503.			
1. AGENCY USE ONLY *(Leave blank)*	**2. REPORT DATE** June 2005	**3. REPORT TYPE AND DATES COVERED** Master's Thesis	
4. TITLE AND SUBTITLE: A Complex Adaptive System Approach to Forecasting Hurricane Tracks			**5. FUNDING NUMBERS**
6. AUTHOR(S) Matthew R. Lear			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Naval Postgraduate School Monterey, CA 93943-5000			**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES) N/A			**10. SPONSORING/MONITORING AGENCY REPORT NUMBER**
11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government.			
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			**12b. DISTRIBUTION CODE**

13. ABSTRACT (maximum 200 words)

Forecast hurricane tracks using a multi-model ensemble that is comprised by linearly combining the individual model forecasts have greatly reduced the average forecast errors when compared to individual dynamic model forecast errors. In this experiment, a multi-agent system, the Tropical Agent Forecaster (TAF), is created to fashion a 'smart' ensemble forecast. The TAF uses autonomous agents to assess the historical performance of individual models and model combinations, called predictors, and weights them based on their average error compared to the best track information. Agents continually monitor themselves and determine which predictors, for the life of the storm, perform the best in terms of the distance between forecast and best-track positions. A TAF forecast is developed using a linear combination of the highest weighted predictors. When applied to the 2004 Atlantic hurricane season, the TAF system with a requirement to contain a minimum of three predictors, consistently outperformed, although not statistically significant, the CONU forecast at 72 and 96 hours for a homogeneous data set. At 120 hours, the TAF system significantly decreased the average forecast errors when compared to the CONU. The multi-agent (MAS) system approach opens the door for statistically significant forecast improvement.

14. SUBJECT TERMS Tropical Cyclones, Hurricanes, Complex Adaptive System, Multi-Model Ensemble, multi-agent systems			**15. NUMBER OF PAGES** 51
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. 239-18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

A COMPLEX ADAPTIVE SYSTEM APPROACH TO FORECASTING HURRICANE TRACKS

Matthew R. Lear
Lieutenant Commander, United States Navy
B.S., United Stated Naval Academy, 1993

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN COMPUTER SCIENCE

from the

NAVAL POSTGRADUATE SCHOOL
June 2005

Author: Matthew R. Lear

Approved by: John Hiles
Thesis Advisor

Patrick Harr
Second Reader

Peter Denning
Chairman, Department of Computer Science

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Forecast hurricane tracks using a multi-model ensemble that is comprised by linearly combining the individual model forecasts have greatly reduced the average forecast errors when compared to individual dynamic model forecast errors. In this experiment, a multi-agent system, the Tropical Agent Forecaster (TAF), is created to fashion a 'smart' ensemble forecast. The TAF uses autonomous agents to assess the historical performance of individual models and model combinations, called predictors, and weights them based on their average error compared to the best track information. Agents continually monitor themselves and determine which predictors, for the life of the storm, perform the best in terms of the distance between forecast and best-track positions. A TAF forecast is developed using a linear combination of the highest weighted predictors. When applied to the 2004 Atlantic hurricane season, the TAF system with a requirement to contain a minimum of three predictors, consistently outperformed, although not statistically significant, the CONU forecast at 72 and 96 hours for a homogeneous data set. At 120 hours, the TAF system significantly decreased the average forecast errors when compared to the CONU. The multi-agent system (MAS) approach opens the door for statistically significant forecast improvement.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

I would first like to acknowledge my family. Their tremendous support has made achieving this goal possible. My wife for understanding the long nights and providing the encouragement when needed. My kids for keeping me laughing and focused on what is really important.

Secondly, I would like to thank fellow student, Jon Wintrode. Jon was always there to answer my questions and teach me the ins and outs of the Java programming language. I could not have done this without his help.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. OBJECTIVE

In the Chief of Naval Operations vision, "Sea Power 21: Projecting Decisive Joint Capabilities," Admiral Clark lays out the three fundamental concepts required for achieving this vision: Sea Strike, Sea Shield, and Sea Basing. Sea Strike is the ability to project offensive firepower for a sustained period throughout the world. Sea Shield ensures defenses are continuously available and Sea Basing is the ability to operate independently on the seas in support of joint forces. Sea Power 21 requires a joint, networked force fed by superior information in order to gain a tactical advantage (Clark 2002). Under the CNO's vision of optimizing the world's largest maneuvering area, the seas, it is essential all meteorological events be accurately predicted to allow for planners to optimally place their assets to exploit the operating environment.

The ability to accurately predict the path and intensity of hurricanes will provide Navy decision makers with superior information to determine the best placement for naval assets. In recent years, the use of artificial intelligence has become more prevalent during the current time of decreasing budgets and manpower. The ability to model events that mimic real life scenarios saves the Department of Defense (DoD) millions of dollars annually. While most DoD ventures into artificial intelligence deal with war-gaming, this experiment will try and use a type of artificial intelligence, adaptive software, to improve hurricane track forecasting.

The objective of this study is two fold. The first objective is to create a hurricane forecast that will produce smaller errors than a consensus forecast of dynamical models. The second objective is to prove an adaptive system is capable of providing the forecaster an objective prediction of a hurricane's path based on a weighted comparison of the adaptive system's decisions and ground truth.

B. MOTIVATION

The ability to reduce position and intensity errors for hurricane forecasting is a vital issue to the United States Navy. During the 2004 Atlantic Hurricane Season, hurricanes caused $45 billion in devastation. The ability to accurately predict its path and potential landfall region far enough in advance to save lives and infrastructure is of severe importance to the Navy and civilian officials. The cost to sortie the Atlantic Fleet runs into the millions of dollars. Coastal evacuations cost local economies millions in lost revenues and wages. An accurate, early hurricane track forecast is essential for planners to minimize the cost of these storms in both lives and damage.

C. BACKGROUND

During the last decade numerical track prediction models have drastically improved and have become indispensable for operational forecasters. This has led to a large number of available model forecasts that has actually turned into a problem for forecasters. The large spread of future storm positions has led to numerous studies as to which model is performing the best (Weber 2003). Adaptive Software, when applied to historical model data, has the ability to make forecast model selections in real time.

1. Multi-model Ensemble Forecasting

Goerss (2000) has shown that a consensus forecast, created by the linear combination of positions from three dynamic models, outperformed the individual models. To analyze to the Atlantic hurricane season, Goerss used the Navy Operational Global Atmospheric Prediction System (NOGAPS; Hogan and Rosmond 1991), the United Kingdom Meteorological Office global model (UKMO; Cullen 1993), and the Geophysical Fluid Dynamics Laboratory Hurricane Prediction System (GFDL; Kurihara et al. 1993, 1995, 1998). The resulting multi-model ensemble forecast reduced 24, 48, and 72 h errors by 16%, 20%, and 23% respectively. In the same study, Goerss analyzed the 1997 North Pacific tropical cyclones using the NOGAPS, UKMO and the global spectral model

(GSM; Kuma 1996). The ensemble forecast improved forecast errors by 16%, 13%, and 12% at 24, 48, and 72 h. The NOGAPS model underperformed the GSM and UKMO models during the 1997 North Pacific Tropical Cyclone season and raised the questions, as to whether an ensemble based on the UKMO and GSM models would perform better than the three-model ensemble. While not statistically significant, the three-model ensemble consistently outperformed the two-model ensemble.

2. Complex Adaptive Systems

A complex adaptive system is a system whose properties are not fully explained by an understanding of its component parts. Complex systems consist of a large number of mutually interacting and interwoven parts, entities or agents (Wikipedia 2005). Examples of complex adaptive systems are social organizations, economies, traffic, and weather. A complex adaptive system does not just passively respond to events. They actively try and turn whatever happens to their advantage (Waldrop 1992). A CAS operates based on three principles: order is emergent as opposed to predetermine, the system's history is irreversible, and the system's future is often unpredictable (Dooley 1996). The basic elements of a CAS are agents. An agent is a software representation of a decision-making unit. Agents have unique traits or personalities, which guide their performance and adaptability. Their actions are based on internal decision rules that depend on imperfect local information (Koritarov 2004).

3. Personality and Variation

When treating a CAS as a population of agents, we must first assume that all the agents are not the same. Variation among agents is an essential requirement for complex, adaptive behavior in a multi-agent system (Axelrod and Cohen, 1999). Initial random sets of predictors and unique personalities for each agent create this type of variation in the TAF program.

The concept of personality in a multi-agent system was first illustrated by the Irreducible Semi-Autonomous Adaptive Combat (ISAAC) multi-agent system for land combat (Ilachinski 1997). This model takes a bottom-up approach to modeling combat that allows for emergent phenomena resulting from nonlinear, decentralized interactions among combatants. The end result of these interactions is a move away from looking at the typical 'equilibrium' solutions among a set of pre-defined aggregate variables. Instead, the system focuses on understanding emergent patterns that surface when a system is out of equilibrium (Ilachinski 1997). The concept of personalities was applied to the El Farol Bar Problem during the MV4015 class, Agent-Based Autonomous Behavior for Simulations (Hiles 2004). This same approach was applied in the TAF program.

4. The El Farol Problem

The TAF program uses sensory input, from an external source, about the location of a storm as feedback ("ground truth") to guide an agent's evolution of its focal predictors. In 1994 Brian Arthur illustrated use of feedback in a multi-agent system in connection with a teaching problem he call El Farol (Arthur 1994).

The idea for this experiment was based on 'El Farol Bar' problem. In this problem, a group of agents must decide whether to go to bar each Thursday night to listen to live music. All agents like to go to the bar unless it is too crowded, that is if more than 60% of the agents go. Each agent is armed with a set of local predictors to help them determine if they should go to the bar. In this case, a predictor might be the average attendance for the past four weeks, the best performing agent's focal predictor, or simply last week's attendance number. Each agent is randomly given a personality of extrovert, introvert, or neutral. The measure of how an agent changes is determined by their personality. For example, if the bar were over crowded one night, the extrovert would decrease its fitness by -1. However an introvert would decrease its fitness by -3, since an introvert personality does not like large social situations. The fitness of an agent

is a numerical assessment of how well an agent is performing. Once an agent's fitness level declines to a predetermined level it will switch out predictors in an effort to become fit. The results of the 'El Farol Bar' problem are such that after an initial variability above and below the 60% threshold, the attendance levels out at 60%. This is a classic example of agents being able to transform their composition to achieve a happy outcome.

5. Hypothesis

The hypothesis for this experiment is that a multi-agent system, based on the principles from the El Farol problem, can create a 'smart' ensemble forecast that will have less error than the consensus forecast as defined by Goerss (2000). One of the primary reasons multi-agent systems have the ability to model complexity is due to the fact they can change their structure based on feedback (Arthur 1994). Chapter II will discuss the data used and the system design of the Tropical Agent Forecaster (TAF) program. Included in this will be a break down of the responsibilities of each major section in the TAF. The analysis of results for the 2004 Atlantic hurricane season will be covered in Chapter III. This will include a comparison of the TAF program results and the consensus forecast results. Chapter IV will define the conclusions and future work possibilities to further enhance the complex adaptive system approach to forecasting hurricanes.

THIS PAGE INTENTIONALLY LEFT BLANK

II. METHODOLOGY

A. DATA

For this experiment, the Automated Tropical Cyclone Forecast Systems (ATCF, Sampson and Schrader 2000) output files for the 2004 Atlantic hurricane season were used to define forecast positions from the suite of numerical models used at the National Hurricane Center (NHC). Specifically, the interpolated versions of the previously mentioned NOGAPS (NGPI), UKMO (UKMI), GFDL (GFDI) as well as the National Center for Environmental Prediction Aviation global model [NCEP AVN (AVNI); Surgi et al. 1998; Lord 1991] and the Geophysical Fluid Dynamics Laboratory – Navy Model (GFNI) were used. Each storm data file contains all forecasts, in 12-hour increments from 00 – 120 h, for the different models. The verifying data, in 6 h positions, are the best-track files pulled from the ATCF. A hurricane is identified if it has a wind speed of greater than or equal to 25 knots. In order to provide more feedback to the agents, the individual dynamic models were interpolated into 6 h forecasts using a simple linear average. The starting date-time-group (DTG) for each storm is determined by finding a common DTG for all five models. The ending DTG, for this experiment, is set by the last available forecast from the NGPI model.

B. ERROR CALCULATIONS

The distance in nautical miles between the verifying position and the forecast position defines the measure of how well the system performs. The forecast position error for model *i*, E_i, is defined to be

$$E_i = \sqrt{(C_i^2 + A_i^2)}, \quad (1)$$

where C_i and A_i are the across track and along track errors, respectively (Goerss 2000). For this experiment we are not concerned with whether the position lags or leads the best track position. Speed and direction are not part of determining how well a predictor or forecast performs.

C. SYSTEM DESIGN

The Tropical Agent Forecaster program is written using the object-oriented Java programming language. An overview of the agent forecasting process is presented in Figure 1. Each agent uses its own set of predictors (its Focal Predictors) to compare with the ground truth (best track data) and adjusts the weights of the Focal Predictors according to their performance, as measure against ground truth. All agents report their Active Predictor (the highest weighted Focal Predictors) for each forecast time to the Tropical Agent Forecaster. After processing all the predictors from the agents, the Tropical Agent Forecaster generates an active forecast by averaging the best predictors for a given forecast time. The basic structure and information flow of the program (Figure 2) is contained in three levels, defined as the predictors, the agents, and the tropical agent forecaster.

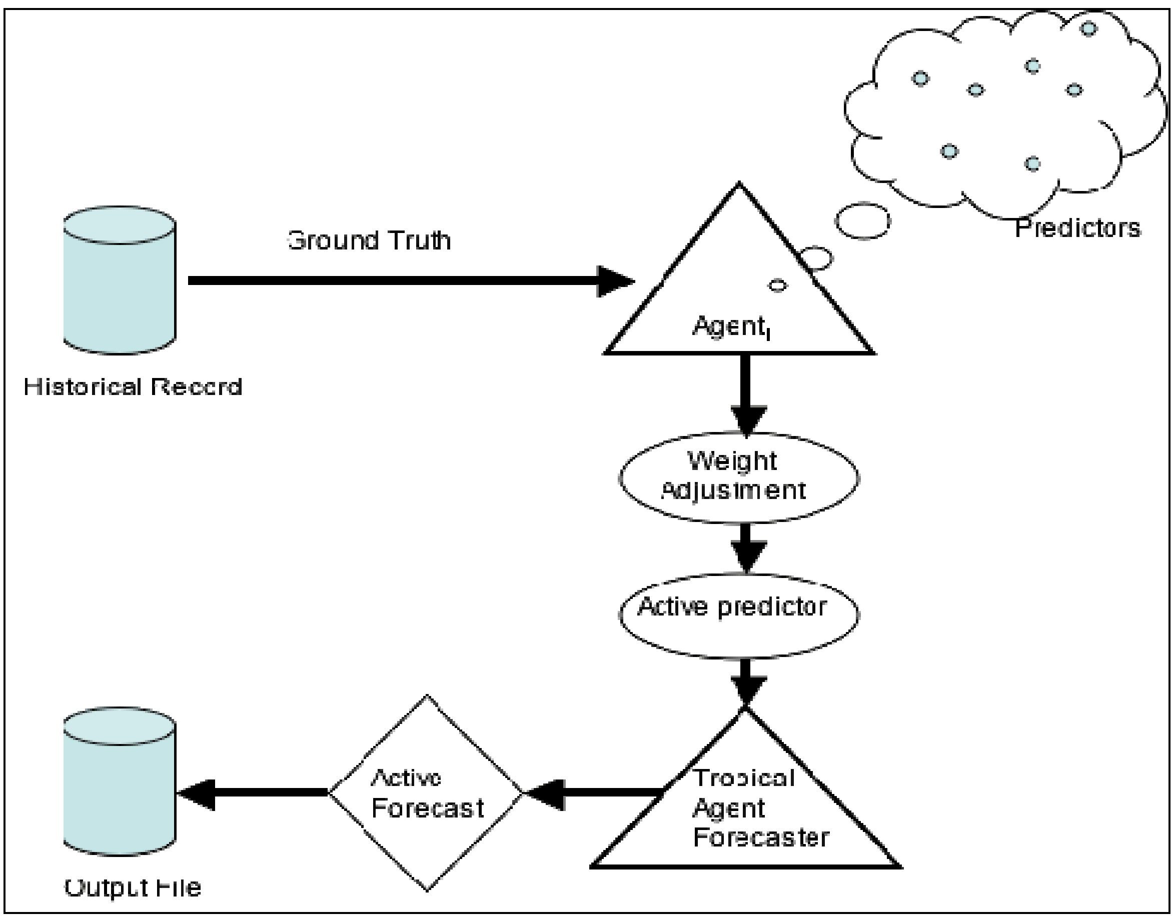

Figure 1. Agent Forecasting Process

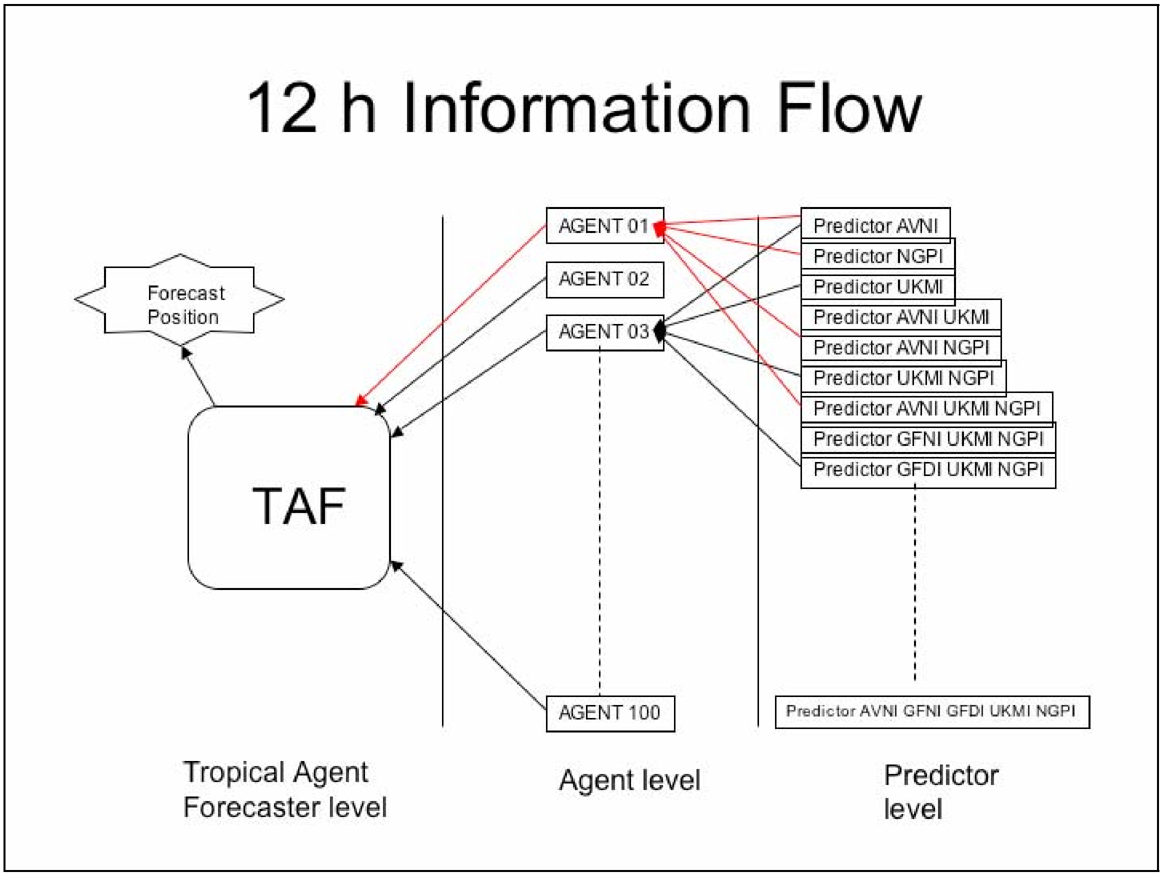

Figure 2. TAF levels and Information Flow

1. Predictors Level

The predictors level contains all the possible combinations of the five dynamic models. Each model combination is a separate predictor and is available for each forecast time (6, 12, 18, 24, 30, 36, 42, 48, 54, 60, 66, 72, 96, 120 h). The two functions of a predictor are 1) get a historical position given a DTG and a forecast time and 2) when directed, get a forecast position for a future DTG and forecast time. An example of a predictor is the UKMI NGPI AVNI predictor. This predictor is a linear combination of positions from each of the specified models at a given forecast time. For this predictor to be created, all three models must be available for the given DTG and forecast time. If one model is not available, the position is set to latitude 0.0, longitude 0.0. This will result in high error numbers and the combination will not be used for the current

DTG. Subsequently, there are several other predictors that can come from this combination, such as an AVNI NGPI predictor, a NGPI UKMI predictor, an AVNI predictor, etc. All possible combinations of predictors are evaluated in 6 h increments and for each possible forecast time.

2. Agent Level

The building blocks of any complex adaptive system are the agents. From a programming point of view, agents are active objects that have been defined to simulate parts of a model (Amin and Ballard 2000). Agents have the ability to evolve in response to their environment. In our program, agents are random given a set of eight predictors for each possible forecast time. That is to say a set of eight 6 hour predictors is randomly assigned, a set of eight 12-hourour predictors is randomly assigned, etc., until all forecast times have been included. The predictor sets for 6 and 12 hours will not be the same. In the end, each agent will have fourteen sets of eight predictors.

The item that differentiates one agent's behavior from the other is their personality. In our program, an agent is either tolerant or intolerant of error. Each agent will weight its local predictors based on their personality. A tolerant agent will react slower to under performing predictors, while an intolerant agent will want to quickly swap out predictors that are underperforming. An example of how error tolerance differs between the two personalities for a 12-hour prediction is provided in figure 3. The effect is to place a target over the current position of the hurricane. The agent, for 12-hour predictors, will look back 12 hours and get the 12-hour forecast for each local predictor. This 12-hour forecast is valid for the current DTG. The intolerant agent will assign a +4, 0, -4, -8 weight to a local predictor if its 12-hour forecast position falls within 0 - 30 nm, 31 – 45 nm, 46 – 60 nm, > 60 respectively. A tolerant agent will assign a +4, 0, -4, -8 weight to a local predictor if its 12-hour forecast position falls within 0 – 42 nm, 43 – 60 nm, 61 – 90 nm, > 90nm respectively. The 12-hour radius for the intolerant agent was set just below the 12-hour total average error of the five models during the season.

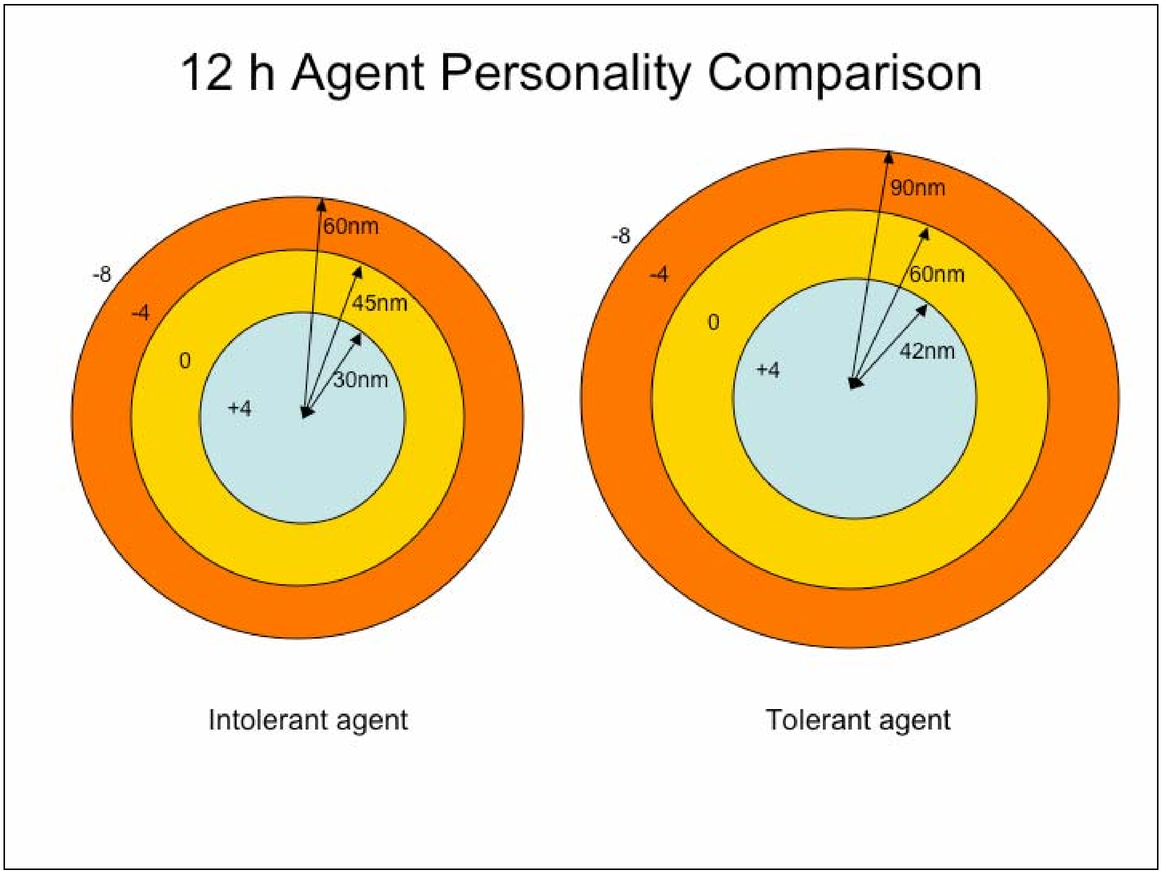

Figure 3. A 12-hour Agent Personality Comparison

Agents have the ability to swap out predictors once the predictor's weight has fallen below a designated fitness value. For this experiment, the fitness value has been set at -12. After each iteration through the forecast cycle, the agent checks the local weights of its predictors. If a predictor has a weight that is below the fitness value, the agent will request a new predictor. This new predictor is guaranteed not to be the same predictor that was just swapped out. This new predictor comes into the agent's set of predictors with a weight of 0.

Once an agent has assessed the performance of each set of its predictors, the agent must designate its best performing predictor. The best performing predictor is the one with the highest weight. If more than one predictor has the same weight, one is chosen randomly from the evenly weighted predictors. The best performing predictor for each forecast time per agent is

made available to tropical agent forecaster level. For each DTG an agent will present 14 predictors, one for each forecast time, to the tropical agent forecaster.

3. Tropical Agent Forecaster Level

The tropical agent forecaster (TAF) is responsible for generating the official forecast for the system. The TAF polls the different agents for their best predictors for each forecast time. Much like how it is done within each agent, the TAF selects the predictors for each forecast time with the highest weight. More often than not, there is more than one predictor with the same weight. This is where the TAF predictor selection differs from the agents. The TAF does not randomly pick one predictor, but rather it simply eliminates duplicate predictors. What is left is a set of equally weighted, unique predictors. The TAF then gets a forecast position for each predictor in the set. To output only one forecast position, the TAF performs a linear average of the highest weighted forecast predictors.

4. Program Information Flow

Upon program initialization, the user selects the storm to analyze. Once the storm has been selected, the ATCF data fields for that storm are loaded and the model data is interpolated into 6 h increments. After data have been ingested, the agents are created. Each agent is randomly given a personality and a set of 8 predictors for each forecast time. Now that each agent has all the information it needs, it begins processing the data fields.

The TAF, like any other CAS, needs a history in order learn and make forecasts. Since at the start of storm there is no history available, the program must wait 6 hours until it can look back 6 hours to assess performance. After 6 hours, the agents will process their set of 6-hour predictors. A 6-hour predictor will look back 6 hours and get its 6-hour forecast. This forecast will be compared with the current position of the hurricane and the error will be calculated. Once the errors have been calculated for each of the 6-hour predictors, each agent will

adjust their local predictor weights based on performance. The agents then check their fitness level and if it is below a threshold of -12, it will swap out its worst performing predictor. Each agent passes the best predictor to the tropical agent forecaster. Once the tropical agent forecaster has each agent's 6-hour prediction, it finds the highest weighted predictors and eliminates duplicates. With this final set of best predictors, the agent gets the 6-hour forecast position from each predictor. These positions are averaged to produce the final 6-hour forecast position. The tropical agent forecaster calculates the forecast error from the best track data and writes this information to a forecast file.

This process is repeated until it reaches the ending DTG. On the second time through the loop, the program is now 12 hours into its analysis. The 6-hour predictors are processed again and now the 12 hours begin to be processed (Figure 4). The process of getting a 12-hour forecast involves going back in time to process the predictors, assigning weights to predictors, and generating a forecast based on the highest weighted predictors. The end result after 12 hours is both a 6-hour and a 12-hour forecast. Every 6-hours another set of predictors is introduced into the system and another forecast is added. The program will generate forecasts in 6-hour increments up to 72 hours and then it generates 96- and 120- hour forecasts.

What makes this forecast position unique to any other multi-model ensemble is the different models and model combinations used to generate the position.

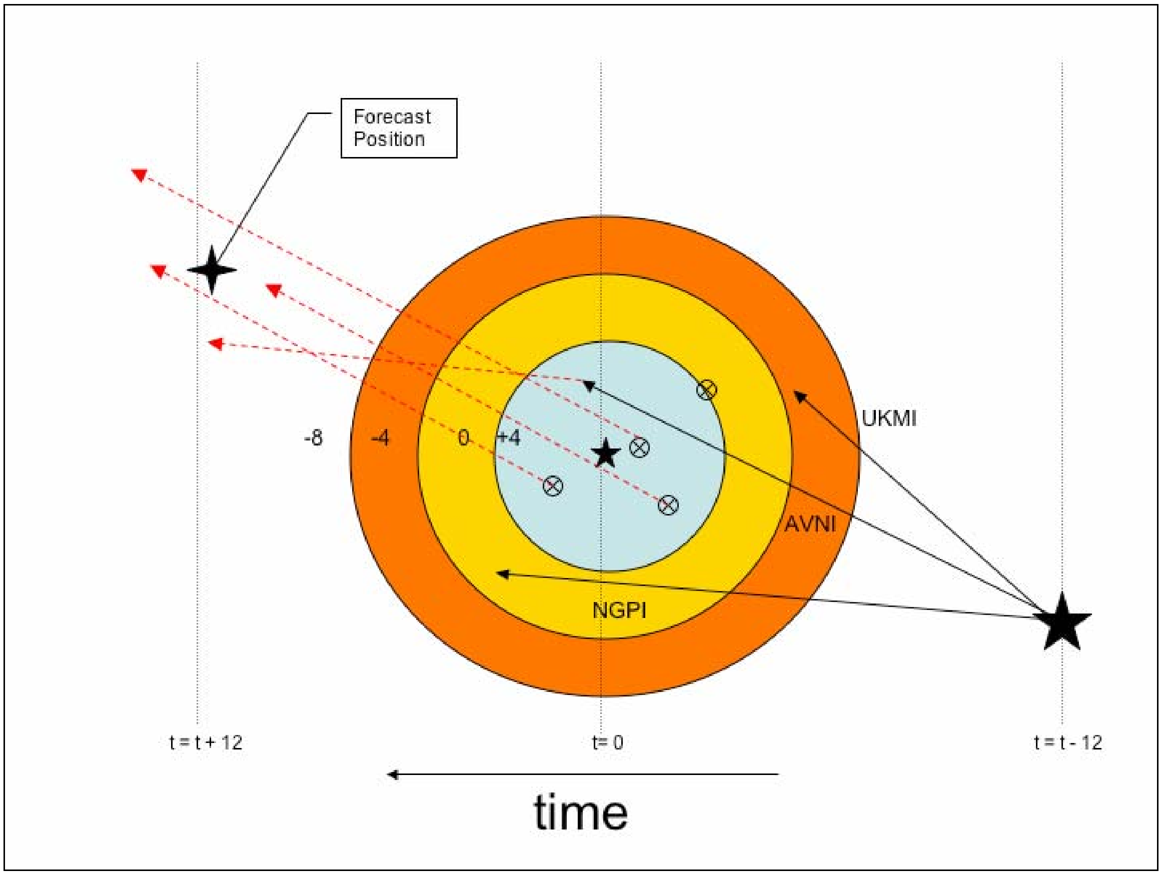

Figure 4. A 12-hour forecast example – the stars represent best track positions, the circles with x inside indicate average positions between models. The four-pointed star is the final forecast position after averaging the forecast positions of the highest weighted predictors.

A final forecast position that is based on an average of the AVNI forecast, the UKMI NGPI forecast, the AVNI NGPI forecast, and the AVNI UKMI NGPI forecast is presented in Figure 4. Below is a sample of the typical output for a 12-hour forecast position.

20048212,

12,

32.2, 77.9,

19.377499622275813,

AVNI UKMI GFDI 12 hour predictor, AVNI UKMI NGPI GFDI 12 hour predictor, AVNI 12 hour predictor, GFDI 12 hour predictor, AVNI and GFDI 12 hour predictor, AVNI UKMI NGPI 12 hour predictor, AVNI and NGPI 12 hour predictor,

The first line is the current DTG. In this case it is August 2, 2004 at 12 Z. The second line indicates this is a 12-hour forecast, and the third line gives the forecast position for August 3, 2004 at 00Z. The fourth line indicates the error associated with the forecast. The last group of lines shows all the models/ model combinations that went into generating the final forecast position.

D. CONSENSUS FORECASTS

The goal of this experiment is for the TAF program's forecasts errors to be significantly less than those of the consensus forecast (CONU). The CONU forecast is a linear combination of individual model forecast positions. The CONU forecast used for comparison in this experiment is comprised of the AVNI, the GFDI, the GFNI, NGPI, and UKMI models. Goerss (2000) showed that a CONU forecast containing three models (UKMO, GFDL, NOGAPS) outperformed individual models throughout the course of the 1995-96 Atlantic hurricane seasons. In a study of the 1997 North Pacific tropical cyclones, the three-model consensus forecast again beat the individual model forecasts. In this case, two of the three individual models significantly out performed the third model. This led to the question of whether a two-model consensus forecast would produce better results. Despite the better individual performance of the two models, the three-model consensus forecast consistently outperformed, but not statistically significant, the two-model consensus forecast (Goerss 2000). The determination was made that a consensus forecast should contain a minimum of three numerical models.

E. COMPARISON OF CONU AND TAF STRUCTURE

There is a significant difference in the manner the TAF system selects its forecast and the manner the CONU calculates its forecast. The CONU forecast

is based simply on model output. That is to say, if all models are available, then a forecast is produced. If one model is not available, then the CONU forecast cannot be calculated. The TAF system is an automatic forecast system. As long as it is feed data it will continue to forecast. The data does not have to consist of all numerical models. It will function on a reduced set of models. The CONU produces a single answer each time it executes, where as the TAF output is produced by the multi-agent system, as long as it receives ground truth and model input. The CONU is an algorithm that works on data. The multi-agent system modifies itself based on feedback from the ground truth. The TAF changes itself through its built in mechanism for obtaining new predictors. If an agent's forecasts perform poorly, then the agent will replace its worst predictor with a new predictor.

III. RESULTS AND ANALYSIS

A. RESULTS

A homogeneous comparison of the hurricane track performance of the NGPI, GFDI, GFNI, AVNI, UKMI, CONU, and TAF is presented in Table 1 for the 2004 Atlantic hurricane season.

	12h	24h	36h	48h	72h	96h	120h
NGPI	39.8	73.2	101.1	137.6	219.2	271.0	375.8
GFNI	41.6	77.5	107.5	156.8	209.3	228.6	416.1
AVNI	38.8	69.3	98.6	147.8	180.1	171.7	291.9
UKMI	40.9	68.9	90.4	124.6	164.4	250.2	234.4
GFDI	34.8	63.2	91.0	140.0	169.4	236.1	279.6
CONU	33.9	61.4	82.8	122.5	152.5	169.3	270.0
TAF	34.8	59.6	87.6	137.9	166.6	190.9	249.4
CASES	186	160	143	113	66	38	20

Table 1. Total errors (nm) for 2004 Atlantic hurricane season

Hurricane forecast errors for the five models and the consensus ensemble were gathered using software from the ATCF system. The TAF forecast errors were output from the program described in chapter II. A Student t-test (Wilks 1993) was performed to assess the statistical significance between the errors associated with the TAF and CONU forecasts. At 12 and 24 hours, the differences between the TAF forecasts and the CONU forecasts are not statistically significant. The TAF program performed significantly worse at 36, 48, and 72 hours. At 96 hours, the TAF program was outperformed by the CONU right at the 95% level, while at 120 hours the TAF program performance was significantly better than the CONU. The remaining analysis will focus on 72 – 120 hours since the ensemble forecasts are most beneficial at the longer forecast intervals where the spread between models tends to increase.

Based on Table 1, it was necessary to examine the individual forecasts preferred by the TAF to answer why its forecast errors were greater than the CONU errors. Initially, the number of times the TAF program gave a better forecast compared to the CONU was identified (Table 2). At 72 hours, the TAF

program gave a better forecast than the CONU 44% of the time. This percentage increased at 96 and 120 hours to 54% and 64% respectively. Because of the TAF design and use of predictors by the agents, it is possible that a TAF forecast may be based on a single model or a combination of models; For example, a 96-hour forecast may be the 96-hour forecast for the AVNI. This would occur when the AVNI has been performing accurately in the past positions such that it is be assigned a high weight.

Based on the results in Table 1, the number of TAF forecasts that are based on a single model is defined and compared with the number of times TAF forecast positions are based on combinations of model positions (e.g. NGPI AVNI). Identification of TAF forecasts based on a single model revealed that the differences between the CONU and single model-based TAF forecasts were large. Using the number of times the TAF program selected only one model, the single model forecast positions data were removed from both the TAF and CONU output and the average errors were recalculated on the new homogeneous set (Table 3). Both forecasts improved at 72 hours, however the improvement of the TAF program was significantly better. At 96 hours, the TAF program went from performing significantly worse to outperforming the CONU, however not at a significant level.

	BETTER			WORSE		
	ALL	SINGLE MODEL	COMBO	ALL	SINGLE MODEL	COMBO
72 h	29	1	28	36	4	32
96 h	21	1	20	17	4	13
120 h	13	2	11	7	3	4

Table 2. Comparison of time TAF program was better of worse than CONU for 72 – 120 h. Also indicated is the number of times individual models were chosen by the TAF program.

The improvement at 96 hours for the TAF program after removing the single model selections is significant. Both the models performed worse at 120 hours. The degradation of 120-hour error is due to Hurricane Frances such that the TAF

program rated the AVNI 120 hour predictor as the best predictor and used it for every 120-hour forecast. Early on in the lifetime of Frances, this AVNI 120 forecast positions vastly outperformed the CONU, but for the final three forecasts, the CONU greatly outperformed the TAF's selection of AVNI 120.

SINGLE MODELS INCLUDED

	72 h	96 h	120 h
CONU	152.5	169.3	270.0
TAF	166.6	190.9	249.4
CASES	65	38	20

SINGLE MODELS REMOVED

	72 h	96 h	120 h
CONU	143.4	177.2	390.0
TAF	150.1	173.1	350.1
CASES	60	33	15

Table 3. Comparison of average errors for 72, 96, 120 hours with single models included and after removing single model selections

After removing the single models, the standard deviations were greatly reduced for both the CONU and TAF (Table 4). The decrease in standard deviation was most significant at 72 and 96 hours, while the increase at 120 hours was somewhat related to the small sample size.

These results led to the conclusion that single model performance will either greatly outperform or under perform a consensus model forecast and will lead to a higher standard deviation for forecast errors. Therefore, the key is to recognize when a single model is performing well. The TAF approach uses past model performance as a predictor to define when an individual model is performing well. Unfortunately, results in Tables 2 and 3 suggest that past model performance is not always related to future performance. Therefore, the TAF system may choose a single model forecast more often than a combination of models.

STANDARD DEVIATIONS (NM) WITH SINGLE MODELS INCLUDED

	72 h	96 h	120 h
CONU	67	113	131
TAF	83	112	137
CASES	65	38	20

STANDARD DEVIATIONS (NM) WITH SINGLE MODELS REMOVED

	72 h	96 h	120 h
CONU	67	106	89
TAF	78	96	116
CASES	60	33	15

Table 4. Comparison of standard deviations (nm) with single models included and without single models included.

Based on the above analysis, it was decided to investigate the impact on the TAF forecast that results from an increased number of predictors. The number of times the TAF program made a forecast using one predictor, two predictors, or three or more predictors was defined (Table 5). In this case, a single predictor is made up from a combination of more than one model. The TAF program selected a single predictor as it forecast solution 18 times between 72 and 120 hours. A two-predictor forecast is when the final forecast is made up two forecast predictors averaged together. It follows that a forecast based on three or more predictors uses an average of three or more forecast positions. When examining the average error for each of these three categories, the three or more predictor forecast for the TAF program was lower that the CONU model at each of the 72, 96 and 120 hour forecast intervals.

	1 PREDICTOR		2 PREDICTORS		3 OR MORE PREDICTORS	
	CONU	TAF	CONU	TAF	CONU	TAF
72 h	153.4	168	83.7	134.2	149.5	143.7
96 h	180.8	182.4	152.9	180.8	178.3	161.1
120 h	NO CASES	NO CASES	NO CASES	NO CASES	390	350.1
CASES	18	18	5	5	37	37

Table 5. A homogeneous comparison of TAF and CONU forecast errors when the TAF program selected one predictor, two predictors or three predictors to generate the forecast position

The average error for the TAF program decreased with the greater number of predictors averaged to create the forecast. For 120 hours, all forecasts were made with three or more predictors. At 72 and 96 hours, the selection of two predictors occurred only 8% of the time. Based on the information in table 5, the TAF program was modified to force at least three predictors be averaged to create the forecast predictions. This change only affected the tropical agent forecaster level of the program. It did not change the manner in which the agents weighed each predictor. The highest weighted predictor was always one of the predictors used in the final forecast. The tropical agent would look at the next lowest weighted predictors provided by the agents and include them in the final forecast prediction. Examining the 72 – 120 hours average errors for the modified TAF program, showed improved performance versus the CONU model. The average forecast error for the modified TAF decreased from 166.6 nm, 190.9 nm, and 249.4 nm (see Table 1) down to 148.4 nm, 185.5 nm, and 237.3 nm respectively for 72, 96, and 120 hours. A t-test was performed to check the significance, at a 95% confidence level, of these new average forecast errors versus the CONU model. At 72 hours, the TAF average error went from being significantly larger than CONU to smaller than CONU. For 96 hours, the results were the same as before, with a marginally significant difference that favored the CONU over the TAF, however the difference between the average errors was closer. The modified TAF remained significantly better than the CONU at 120 hours. Standard deviations improved slightly at 72 and 96 hours, however it increased slightly at 120 hours.

B. CASE STUDY

Hurricane Ivan is presented as a case study to highlight an example of when the TAF and TAF-3 programs provided a positive result when compared to the CONU forecast. The complete set of forecast tracks for CONU, TAF, and TAF-3 for Hurricane Ivan (Figures 5, 6, and 7 respectively) define a right (eastward) bias throughout the life of the storm. This is an indication that the majority of models are forecasting positions to the right of the actual hurricane track. It is not possible to eliminate the right side bias with the current configuration of the TAF and TAF-3 programs. The goal is to reduce this bias by selecting a predictor that will not include the largest error models.

[Note: Original track figures were color images. For black and white reproductions each track is a different grayscale value that is consistent for the entire forecast track.]

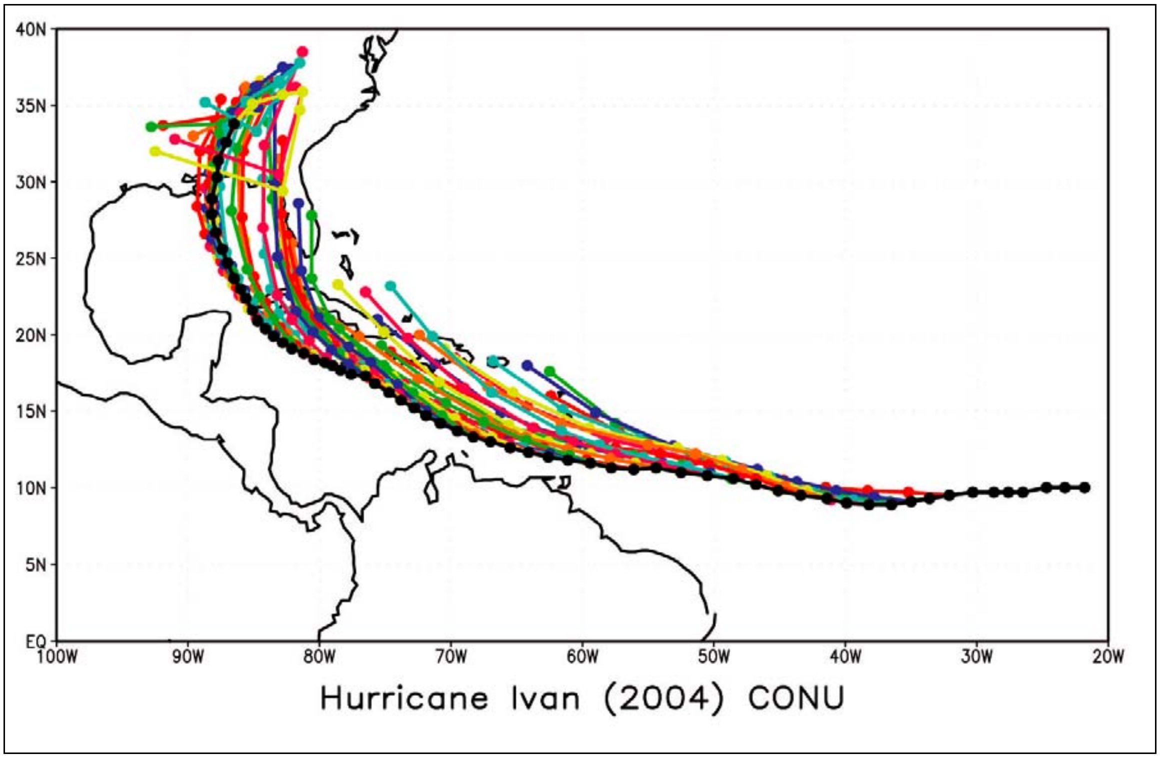

Figure 5. Hurricane Ivan complete set of CONU forecasts. The black circles represent the best track positions in 6-h intervals. Forecast positions are defined by alternating colors at 12-h intervals to 72 hours, then 24-h intervals to 120 hours.

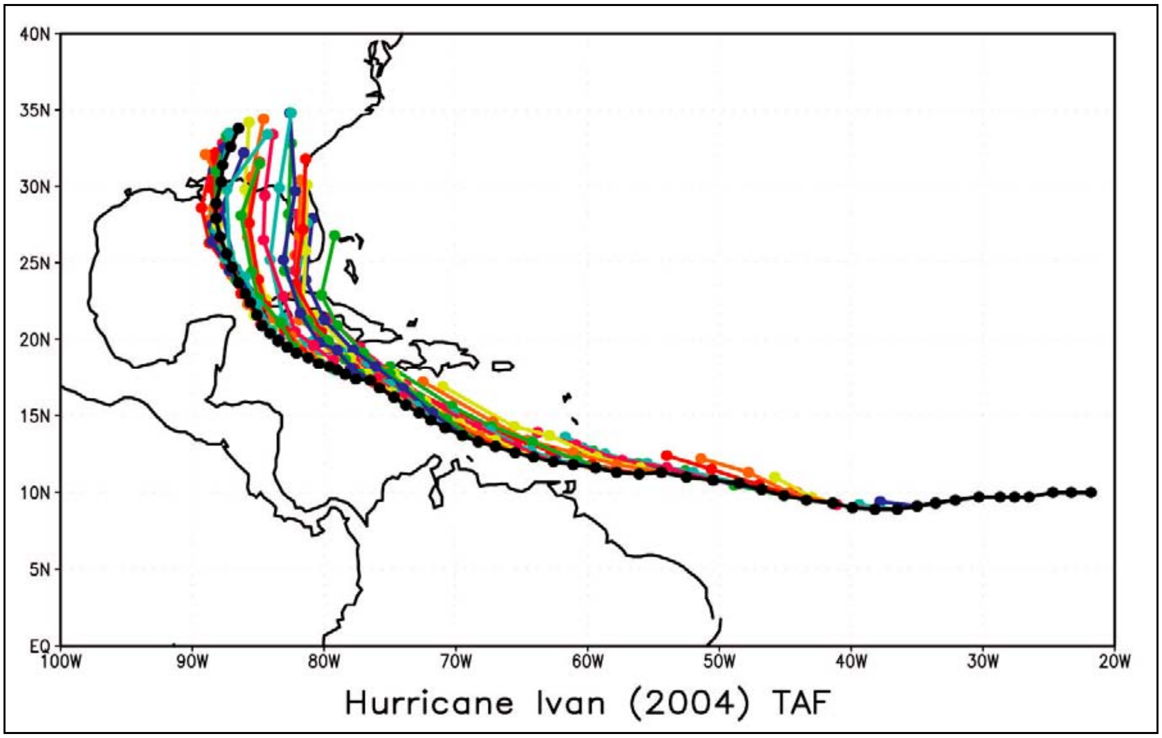

Figure 6. As in Figure 5, except for the TAF forecasts.

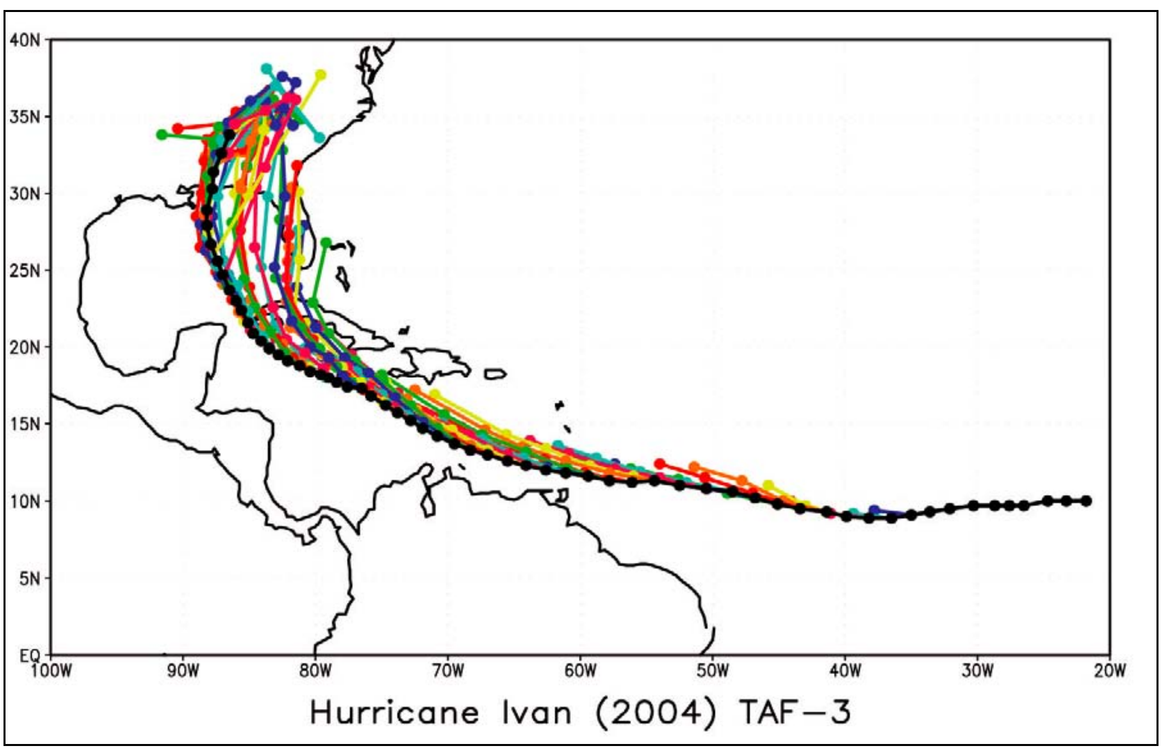

Figure 7. As in Figure 5, except for the TAF-3 forecasts

The 96 and 120-hour errors for Hurricane Ivan indicate that early on in the storm all three forecasts are performing similarly (Figures 8 and 9). The TAF and TAF-3 forecasts errors are slightly less than the consensus forecast error at 1800 UTC on September 8, 2004 (highlighted with the blue rectangle). The green rectangles (Figure 8) highlight the forecast errors at 0000 UTC and 1200 UTC on September 11, 2004 and show that the TAF and TAF-3 96-hour forecast errors (Figure 8) are initially larger than the CONU but the trend is reversed just twelve hours later. At 120 hours (Figure 9) a similar trend is noticed such that the performance of the TAF and TAF-3 become significantly better than the performance of the CONU.

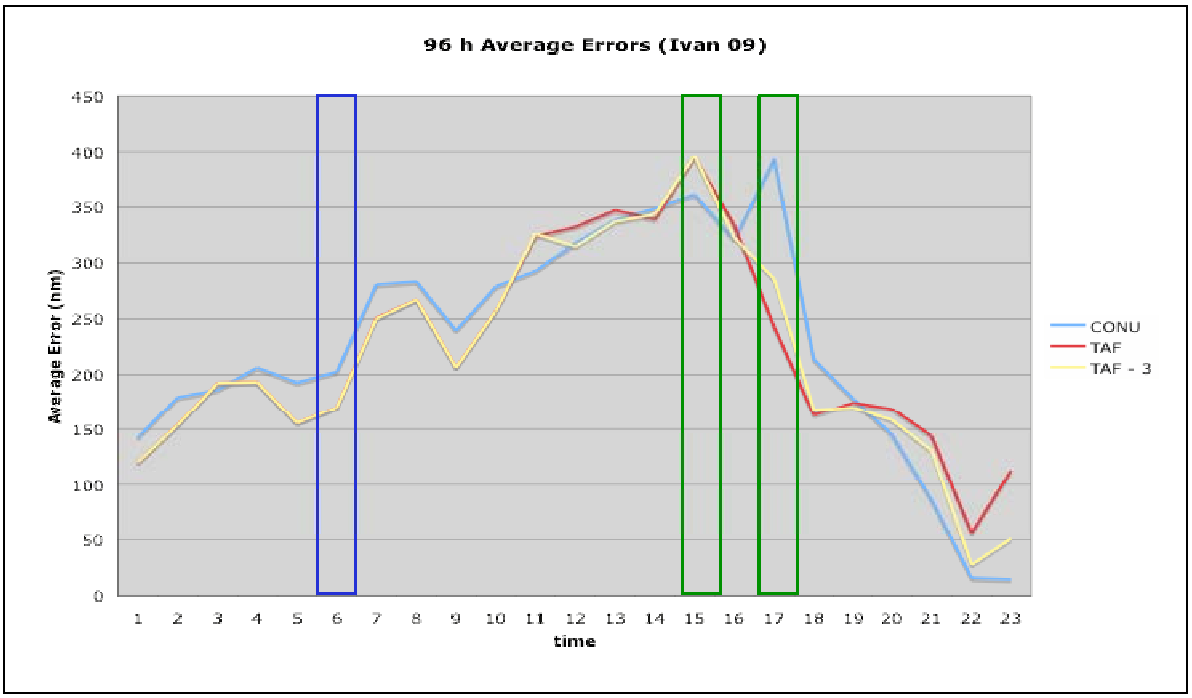

Figure 8. CONU, TAF and TAF-3 96 h forecast errors for Hurricane Ivan

Figure 9. CONU, TAF and TAF-3 120 h forecast errors for Hurricane Ivan

When inspecting the forecast tracks that correspond to the highlighted areas, the type of performance characteristics discussed above become evident. At 1800 UTC on September 8, 2004, all three forecasts are performing in a similar fashion as Hurricane Ivan heads into the Caribbean Sea (Figure 10). At 0000 UTC on September 11, all three forecasts for 96 and 120 hours are moving the storm at a similar speed and there is an insignificant error difference favoring the CONU (Figure 11). Stepping forward to 1200 UTC on September 11 (Figure 12), both the 96 and 120-hour forecasts for the TAF and TAF-3 are significantly outperforming the CONU forecasts. The CONU has accelerated the storm northward much quicker than the TAF and TAF-3. This is caused by the requirement that the CONU contain all models in creating its forecast position. In this case, the NGPI 120 hour error was over 1300nm. This drastically affected the final position for the CONU. The TAF and TAF-3 did not accelerate the storm since the NGPI was not included in any of the predictors used to make its 96 and 120-hour forecast.

Therefore, this example illustrated the ability of the TAF system to recognize that a model is performing poorly and removes it as a predictor for future positions.

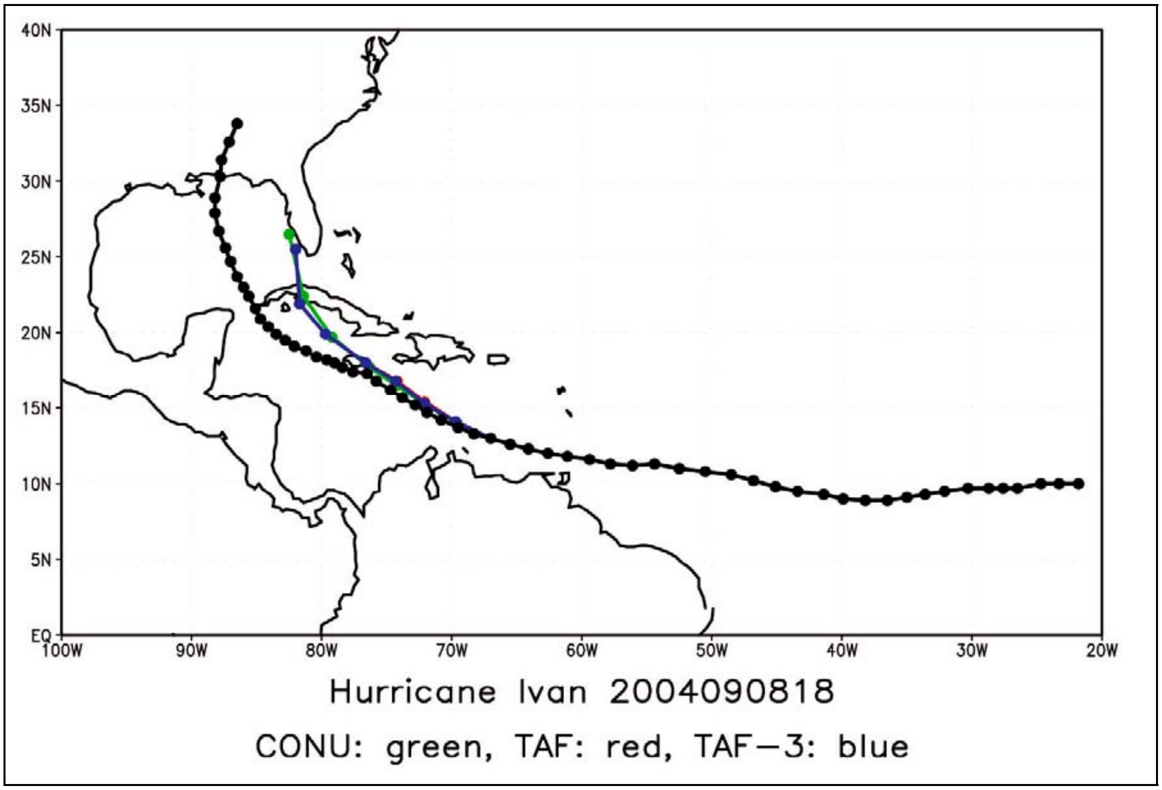

Figure 10. The color scheme is as in Figure 5. This figure shows the forecast (2004090818) tracks for CONU, TAF and TAF-3

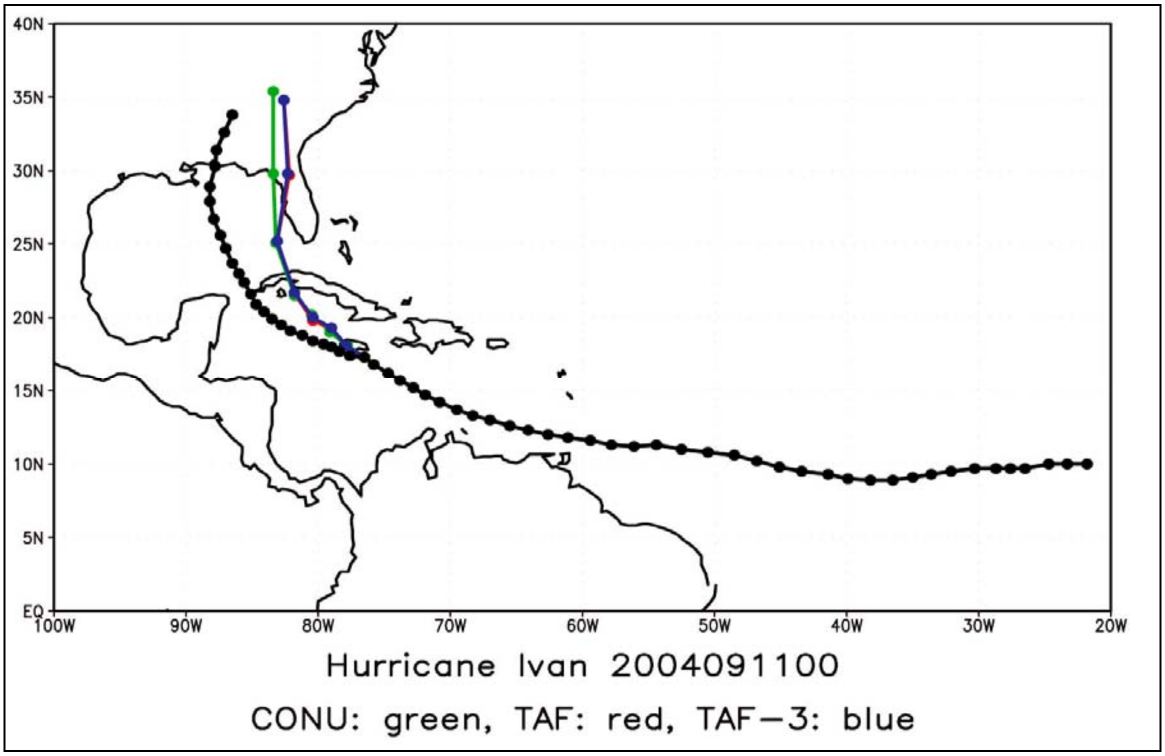

Figure 11. The color scheme is as in Figure 5. This figure shows the forecast (2004091100) tracks for CONU, TAF, and TAF-3

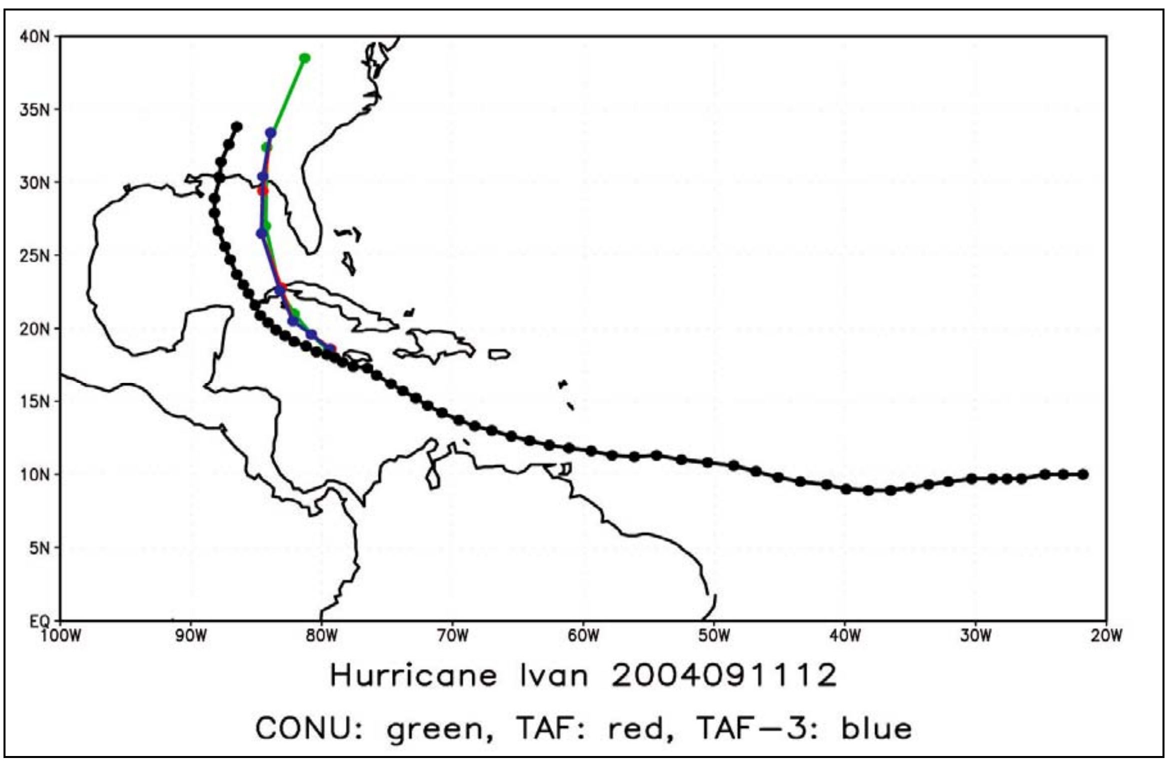

Figure 12. The color scheme is as in Figure 5. This figure shows the forecast (2004091112) tracks for CONU, TAF, and TAF-3

THIS PAGE INTENTIONALLY LEFT BLANK

IV. CONCLUSIONS

A. SUMMARY

A complex adaptive system was created to forecast hurricane track position for the 2004 Atlantic hurricane season. The TAF program used intelligent agents to create a ‘smart’ ensemble based on the historical performance of both individual and combinations of dynamic models. In the initial application of the TAF system, an unconstrained application of the TAF was used such that the absolute set of highest weighted set of predictors was used to produce a forecast position. Based on the TAF design, a predictor may be comprised of a single model or a combination of models. A single model may be the highest weighted predictor when it has been consistently producing highly accurate forecasts over the past lifetimes of the hurricane. Results using the unconstrained system indicated that the TAF forecast were only statistically better than a pure linear combination of all input models at 120 hours.

These results were examined to identify whether the use of single-model predictors caused the TAF to have increased errors. Indeed, removal of single-model based forecast improved the TAF forecast with respect to the linear average of all models. Furthermore, the standard deviation of forecast errors was greatly reduced when single-model forecasts were removed. This is anticipated since the remaining predictors are based on a combination of forecast models.

The final analysis investigated the impact on forecast accuracy from using increased numbers of combination-based predictors. The TAF program, when forced to use three or more predictors, consistently outperformed the CONU forecasts for 72 hours, but the difference was not statistically significant. At 96 hours, the CONU still out performed the TAF program, however the average error difference decreased. There is a statistically significant performance improvement at 120 hours.

However, the ability to use a CAS to predict hurricane tracks has validity. The application of an unconstrained system may need further examination. One note of caution is that the continual average of forecast predictors into one final forecast position decreases the importance of the agents. In an ideal CAS application, one agent's prediction should provide the answer, not a combination of several agent predictions. However, this may be adversely impacted by the fact that, with respect to hurricane track forecasting, past model performance is not significantly correlated to future performance.

For each forecast period, parallel exploration of the problem by the 100 first level agents, each of which exhibit variation, produces a set of autonomous decisions – the 100 Active Predictors. In turn, the Tropical Agent Forecaster uses the active predictors to produce a forecast for that time period. There is a connection between the autonomous decisions from the MAS software system and the ground truth measures of the real world storm positions. The thesis results show that the hypothesis was partial proven. Although, short-range forecasts were not significantly improved, the 120-hour forecast by TAF showed statistically significant improvement. Long-range forecast are particularly important for issuing warnings and evacuation notices. This thesis justifies further exploration of multi-agent and complex adaptive system techniques in connection with hurricane forecasting. This could lead to a reduction in both property damage and loss of life caused by these powerful storms.

B. FUTURE WORK

There are a number of ways to implement a complex adaptive system to forecast hurricane tracks. The current TAF system is a first step in creating an agent based forecasting system. Based on this approach, the following recommendations are provided to improve the application of a complex adaptive system to hurricane track forecasting.

1. Remove Agent Restrictions

The agents in the TAF system are currently given a set of eight predictors segmented into fourteen forecast times. The next generation of TAF should remove the segmentation of the forecast time slots [i.e. 6, 12, 18…]. An agent should be given a total of eight predictors to forecast with for fourteen forecast periods. This would enable the best 6-hour predictor to compete as the best 48-hour predictor and enable the agents to make decisions based on both the best currently performing predictor and the best historical performing predictor. The process of looking back 96 hours to get the best prediction to forecast out 96 hours would be reduced. This will hopefully lead to a more accurate prediction of the future forecast.

2. Creating History

For a complex adaptive system to work it must have an accurate history. For example, the current TAF system must wait 96 hours into the storm's life in order to produce a 96-hour forecast. This reduces the number of long range forecasts to an unacceptably low level, particularly when forecasting in the Atlantic Ocean. It might prove that simply removing the agent restrictions noted above will be sufficient in providing more accurate long-range forecasts. The addition of climatology data into the system might prove useful in creating a history that can be used to forecast longer ranges more accurately from the start of the storm.

3. Pacific Tropical Cyclone Analysis

The TAF program should be used to analyze past Western North Pacific tropical cyclone seasons. The tropical cyclones in the Western North Pacific Ocean usually have longer tracks than those in the Atlantic Ocean. Additional TAF output data collected for the 72, 96, and 120 hour forecast periods would validate the Atlantic Ocean data. Simple data ingest modification that enables the TAF system to recognize which basin the storm is in would make this analysis possible.

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Amin, M., and Ballard, D, 2000: Defining New Markets for Intelligent Agents. *IT Pro.*, **1520-9202/00,** 29 – 34.

Arthur, W. B. 2001: Inductive Reasoning and Bounded Rationality. [Available online at http://www.santafe.edu/arthur/Papers/El_Farol.html], Retrieved June 2005.

Axelrod, R. and Cohen, M., 1999: *Harnessing Complexity,* The Free Press, New York, NY, 184.

Clark, V., 2002: Sea power 21: Projecting joint capabilities, *U.S. Naval Institute Proceedings,* **128,** 32 - 44.

Cullen, M. J. P., 1993: The Unified Forecast/Climate Model. *Meteor. Mag.*, **122**, 81-122.

Dooley, K., cited 2005: A nominal definition of complex adaptive systems. [Available online at http://www.eas.asu.edu/~kdooley/casopdef.html] , Retrieved June 2005.

Goerss, J. S., 2000: Tropical Cyclone Track Forecasts Using an Ensemble of Dynamical Models. *Mon. Wea. Rev.*, **128**, 1187-1193.

Hiles, J., 2004: MV4015 - Agent-Based Autonomous Behavior for Simulations class notes., Naval Postgraduate School, July 2004.

Hogan, T. F., and T. E. Rosmond, 1991: The description of the Navy Operational Global Atmospheric Prediction System's spectral forecast model. *Mon. Wea. Rev.*, **119**, 1786-1815.

Ilachinski, A., 1997: *Irreducible Semi-Autonomous Adaptive Combat (ISAAC): An Artificial-Life Approach to Land Warfare (U).* Alexandria, VA., Center for Naval Analyses.

Koritarov, V. S., 2004: Real-World Market Representation with Agents. *IEEE power and energy magazine.*, **1540-7977/04,** 39 – 46.

Kuma, K., 1996: NWP activities at Japan Meteorological Agency. Preprints, *11th Conf. on Numerical Weather Prediction,* Norfolk, VA, Amer. Meteor. Soc., J15-16.

Kurihara, Y., M. A. Bender, and R. J. Ross, 1993: An initialization scheme of hurricane models by vortex specification. *Mon. Wea. Rev.*, **121,** 2030-2045.

Kurihara, Y., and R. J. Ross, 1995: Improvements in the GFDL hurricane prediction system. *Mon. Wea. Rev.,* **123**, 2791 – 2801.

Kurihara, Y., R. E. Tuleya, and M. A. Bender, 1998: The GFDL hurricane prediction system and its performance in the 1995 hurricane season. *Mon. Wea. Rev.,* **126,** 1306 -1322.

Sampson, C. R. and A. J. Schrader, 2000: The automated tropical cyclone forecasting system (version 3.2)., *Bull. Amer. Meteor. Soc.,* **81,** 1231 – 1240.

Surgi, et. al., 1998: Improvement of the NCEP global model over the tropics: An evaluation of model performance during the 1995 hurricane season., *Mon. Wea. Rev.,* **126,** 1287 – 1306.

Waldrop, M. M. 1992: *Complexity – The Emerging Science at the Edge of Order and Chaos,* Simon and Schuster, New York, NY, 384

Weber, H. C. 2003: Hurricane Track Prediction Using a Statistical Ensemble of Numerical Models., *Mon. Wea. Rev.,* **131,** 749 – 771.

Wilks, D., 1995: *Statistical Methods in the Atmospheric Sciences,* Academic Press, New York, NY, 467.

Wikipedia. The Free Encyclopedia, cited 2005: Complex Adaptive System. [Available online at http://en.wikipedia.org/wiki/Complex_adaptive_system], Retrieved June 2005.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, VA

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, CA

3. Professor Peter Denning
   Naval Postgraduate School
   Monterey, CA

4. Professor John Hiles
   Naval Postgraduate School
   Monterey, CA

5. Professor Chris Darken
   Naval Postgraduate School
   Monterey, CA

6. Professor Rudy Darken
   Naval Postgraduate School
   Monterey, CA

7. Professor Ted Lewis
   Naval Postgraduate School
   Monterey, CA

8. Professor Donald Gaver
   Naval Postgraduate School
   Monterey, CA

9. Professor Pat Jacobs
   Naval Postgraduate School
   Monterey, CA

10. Lieutenant Commander Matthew Lear
    Commander Destroyer Squadron 22
    Norfolk, VA